

# カラー答案画像からの良好な解答文字パターンの分離抽出手法

西 彦樹<sup>†</sup> 松尾 賢一<sup>†</sup>

<sup>†</sup> 奈良高等専門学校 情報工学科

## 1. はじめに

本研究において、答案の採点作業のミスを低減する「採点ミス発見支援システム」の開発に取り組んできた[1]。本システムは、カラー答案画像を入力すると、答案画像内の採点記号と部分点パターンを認識し、その認識結果同士から得点の整合性を判定し、その判定結果から採点ミスの可能性がある箇所を採点者に提示する役割を備えている。現状での本システムは、採点記号と部分点との不一致の有無の存在を採点者に通知する採点ミス発見支援でしかない。本発表では、本システムでの採点ミス発見支援の高性能化に向けて、受験者が鉛筆で答案に記述した解答文字の認識結果も、得点の整合性判定に加えることを目的として、答案画像から良好な解答文字パターンだけを分離抽出する手法を提案する。さらに、提案手法の有効性を解答文字パターンの抽出結果から考察する。

## 2. ラプラシアンヒストグラム(LH)法

LH(Laplacian Histogram)法は、大津の判別分析法[2]の改良手法である。濃淡画像内の対象領域の輪郭付近における濃度分布を調べ、対象領域とそれ以外の背景との濃度に対する濃度ヒストグラムを作成する。このヒストグラムに対して、大津の判別分析法を適用することで、濃淡画像を2値化するための閾値を算出するのがLH法である。このLH法は、背景に対して対象物の面積が比較的少ない解答パターンを抽出する際に有効な2値化手法として用いる。

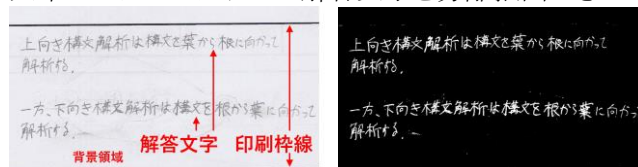
## 3. 提案手法

イメージスキャナ(JustioMFC-6490CN, brother)を用いてA4用紙を解像度300dpiで電子化したカラーの答案画像に対して、従来手法[1]を用いて赤色の採点記号と部分点パターンの両者を除去する。次に、両パターンが除去された画像からラプラシアン<sup>1</sup>の絶対値を算出し、その算出値が一定以上である着目画素周辺において2.のLH法を用いて濃度ヒストグラムを印刷枠線(黒色)、解答文字(灰色)、背景領域(白色)の3つの濃度範囲に分割するための2つの閾値 $Th_1, Th_2$ を算出する。ここで $Th_1$ は、印刷枠線の濃度値と解答文字以上のそれを分離する閾値、 $Th_2$ は、解答文字以下の濃度値と背景領域のそれを分離する閾値となる。この2つの閾値 $Th_1$ と $Th_2$ の間の濃度値が本研究で分離抽出対象としている解答文字パターンの濃度値であることから、この画素

値を1、それ以外を0とした2値画像を得る。この2値画像には印刷枠線の輪郭付近が雑音として発生するため、最後に $Th_1$ 未満の濃度である印刷枠線を輪郭付近まで膨張させたマスク画像を用いて、マスク画像と同じ位置にある枠線輪郭を除去する。

## 4. 分離抽出パターン確認実験

答案画像○枚に対して分離抽出パターンの品質を確認する実験を実施した。図1に本手法による解答文字パターンの分離抽出結果を示す。図1に示すように大半のパターンにおいて解答文字を分離抽出できた。



(a)元画像

(b)分離抽出結果

図1. 解答文字パターンの分離抽出結果

実験結果から提案手法により全パターンに対して一部の画像において、穴や欠陥があるなど良好な分離抽出パターンが得られなかった。特に、これらのパターンでは、LH法での濃度ヒストグラム内で、印刷枠線、解答文字、背景領域の3つのヒストグラムの峰が明確でないため、算出した閾値 $Th_1$ で印刷枠線と解答文字間の濃度値を良好に2分していないことが明らかとなった。

## 5. おわりに

画像全体を1つの閾値で2値画像を生成していることや、解答文字パターンの濃度範囲の一部が印刷枠線のそれと重畳している場合などの要因で、一部の解答文字パターンで欠損が生じていた。これに対して、局所2値化や分離抽出パターンの品質の良好性を考慮した閾値算出手法を導入した提案手法の改善を図っている。(改善結果については当日のポスターで発表)

## 参考文献

[1] 村本ら, “採点ミス発見支援システムにおける分離抽出精度向上に向けた改良”, 情報科学技術フォーラム(FIT: Forum on Information Technology)2022, CH-014, 2022.

[2] 大津展之, “判別および最小2乗基準に基づく自動しきい値選定法”, 信学論(D), Vol.J63-D, No.4, pp.349-356, 1980.

謝辞 本研究はJSPS科研費20K03143の助成を受けたものである。